

## Two phase formation of massive elliptical galaxies : study through cross-correlation including spatial effect

Soumita Modak<sup>1</sup>, Tanuka Chattopadhyay<sup>2</sup>, Asis Kumar Chattopadhyay<sup>3</sup>

<sup>1</sup> Department of Statistics, University of Calcutta, 35 B.C. Road, Kolkata 700019, India; email: soumitamodak2013@gmail.com

<sup>2</sup> Department of Applied Mathematics, University of Calcutta, 92 A.P.C Road, Kolkata 700009, India; email: tanuka@associates.iucaa.in

<sup>3</sup> Department of Statistics, University of Calcutta, 35 B.C. Road, Kolkata 700019, India; email: akcstat@caluniv.ac.in

### Abstract

Formation mechanism of present day population of elliptical galaxies have been revisited in the context of hierarchical cosmological models accompanied by accretion and minor mergers through cross correlation function including spatial effect. The present work investigates the formation and evolution of several components of nearby massive early type galaxies (ETGs) through cross-correlation in the spatial coordinates, right ascension and declination (RA, DEC) and mass-size parameter space with high redshift ( $0.5 \leq z \leq 2.7$ ) ETGs. It is found that innermost components of nearby ETGs are highly correlated with ETGs in the redshift range ( $2 \leq z \leq 2.7$ ) known as 'red nuggets'. The intermediate and outermost parts have moderate correlations with ETGs in the redshift range ( $0.5 \leq z \leq 0.75$ ). The quantitative measures are highly consistent with the two phase formation scenario of massive nearby early type galaxies as suggested by various authors and resolves the conflict raised in a previous work suggesting other possibilities for the formation of outermost part of nearby massive ETGs. The improvement is expected to be due to inclusion of spatial effects in addition to other linear parameters.

keywords: galaxies:formation, galaxies:evolution, methods:statistical

## 1 Introduction

In 1934, Hubble first studied the frequency distribution of galaxies according to source points in the Universe and it was found to be strongly skewed which

indicates that galaxies have spatial clustering nature. The clustering nature is intrinsic compared to uniform distribution (Bok 1934; Mowbray 1938). Numerous studies have been carried out to explore the clustering nature using position co-ordinates (viz. RA and DEC) which are angular in nature (Chandrasekhar & Munch 1952; Zwicky 1953; Limber 1953, 1954; Neyman et al. 1954). In a previous work (De et al. 2014), we have investigated the clustering nature with respect to other intrinsic properties of galaxies, e.g. mass, size, etc. which are linear in nature. In the present work both angular as well as linear parameters are synchronized to study the clustering nature in greater detail.

The discovery of compact massive early type galaxies at high redshift has raised questions on the plausible formation mechanisms for the present day population of elliptical galaxies from the classical view of a single event, e.g. monolithic collapse (Larson 1975; Carlberg 1984; Arimoto & Yoshii 1987) or a major merger (Toomre & Toomre 1972; Ashman & Zepf 1992; Zepf et al. 2000; Bernardi et al. 2011; Prieto et al. 2013). Instead, the single events have to be embedded in hierarchical merging, called minor mergers accompanied by continuous accretion (Forbes et al. 1997; Bluck, Conselice & Buitrago 2012; Newman et al. 2012; Mondal et al. 2008). Recent discovery of compact high redshift galaxies (Daddi et al. 2006; Damjanov et al. 2011; Cappellari et al. 2009; Onodera et al. 2012) and intermediate redshift galaxies having stellar masses and sizes increased by a factor 3 – 4 (van Dokkum et al. 2010; Papovich et al. 2012; Szomoru et al. 2012) suggest a two phase galaxy formation scenario of nearby massive early type galaxies (ETGs) (Forbes et al. 2011; Gobat et al. 2011; Oser et al. 2010; van Dokkum et al. 2015; de la Rosa et al. 2016; Vulcani et al. 2016). These two phases (viz. inside-out) i.e. first an intense dissipational process like accretion (Dekel, Sari & Ceverino 2009) or major merger forms an initially compact inner part and then a slower phase starts when the outermost part is developed through non-dissipational process, e.g. dry merger. This two phase scenario (Oser et al. 2010; Johansson et al. 2012) has severe challenge over classical ‘single event’ model like ‘monolithic collapse’ or ‘major merger’. In a previous work (Huang et al. 2013) the authors have explored the above mentioned possibility through matching ‘median’ values of the two systems. Their results, though support two phase scenario, have the following limitations. They have considered only univariate data like ‘mass’ or ‘size’ at a time. For nearby ETGs, inner, intermediate and outer parts have mass-size correlations with p-values as  $r(M_*, R_e) = 0.817$ , p-value= 0,  $r(M_*, R_e) = 0.623$ , p-value= 0

and  $r(M_*, R_e) = 0.668$ , p-value= 0 respectively. Use of univariate median matching is not suitable for highly correlated bivariate data. Also median does not include complete information in a particular data set. In fact, the mass-size data gave a different result in a previous work (De et al. 2014). Hence for the above study a more sophisticated method should be used. In the above mentioned work the authors have compared the mass-size data of high redshift galaxies with nearby massive ETGs through cross-correlation which is especially designed to study such bivariate data. But this method used only mass-size parameters and not their spatial coordinates (e.g. RA and DEC), whereas spatial coordinates are equally important parameters for description of spatial clustering, if any. In a word, mass-size bivariate data are significant for exploring intrinsic clustering nature and spatial coordinates are significant for spatial clustering. So it will be highly tempting to combine the above two to study the interaction between them.

In the present work we have considered both the intrinsic as well as spatial parameters, i.e. we have studied the cross-correlation among objects having linear as well as angular data. This is a new technique developed for the first time to apply for such multivariate astrophysical data sets. This is more trustworthy and meaningful than that used in the previous works. In Section 2, we have discussed the data sets whereas Section 3 describes the method. The results and interpretations have been structured under Section 4.

## 2 Data sets

In the work by Huang et al. (2013), masses and effective radii of 70 nearby ETGs are given for three parts. (i) Inner component with effective radii  $R_e \leq 1$  kpc, (ii) intermediate component with effective radii  $R_e \sim 2.5$  kpc and (iii) outer envelope with  $R_e \sim 10$  kpc. In the present study these data have been combined with spatial data RA and DEC. In a previous work (De et al. 2014), the above mass-size data for inner, intermediate and outer most parts were stored in three sets and they were called as data sets 1, 2 and 3 respectively. In the present case the situation is different. Since data sets 1 – 3 are the values of the same galaxies, we can not just associate the same RA and DEC in three sets, because in that case RA and DEC will be the same for three data sets and the effects of RA and DEC will not be reflected in the analysis. Hence we have to consider three data sets where all RA, DEC, mass and size have different values. Here RA, DEC are angular

data and mass, size are linear data. It is difficult to apply various statistical methods to such heterogeneous data sets. Hence we have linearized RA and DEC (viz. Subsection 3.1). We have carried out k-medoids cluster analysis using Euclidean distance (Kaufman & Rousseeuw 1990) of 70 galaxies having parameters RA, DEC (linearized), mass and size into three groups for each of the galaxy components (viz. inner, intermediate and outer most parts). Thus each component of nearby ETGs has been divided into three groups. We have chosen data set 1 as the cluster of galaxies having the minimum average size ( $\langle R_{e,1} \rangle$ ) from the innermost component groups, data set 3 as the group having the maximum average size ( $\langle R_{e,3} \rangle$ ) from the outermost component groups and data set 2 as the group having medium average size ( $\langle R_{e,2} \rangle$ ) from the intermediate parts in order to get best performance. Now all the values in data sets 1 – 3 have different values of the parameters for comparison. We have also tried other combinations for data sets 1 – 3 among twenty seven possibilities. The results are more or less similar. In the first choice, data sets 1 – 3 have sizes 15, 25 and 30 respectively.

Data sets 4 – 8 consist of mass-size data of ETGs in the high redshift zone (viz.  $0.5 \leq z \leq 2.7$ ) and their masses have the lower limit  $M_* \geq 10^{8.73} M_\odot$ . The entire redshift zone has been divided into five equal bins,  $0.5 < z \leq 0.75$ ,  $0.75 < z \leq 1$ ,  $1 < z \leq 1.4$ ,  $1.4 < z \leq 2$  and  $2 < z \leq 2.7$  similar to De et al. (2014). The present data set of high redshift ETGs is a larger one from the previous one and the size is 1012 as we have included more recent data in it. RA, DEC, masses and sizes of the ETGs are taken from the following works.

364 ETGs ( $0.19 \leq z \leq 2.67$ ) from Damjanov et al. (2011), 248 ETGs ( $1.49 \leq z \leq 1.79$ ) from Papovich et al. (2012), 142 ETGs ( $0.70 \leq z \leq 1.31$ ) from Chen et al. (2013), 177 ETGs ( $0.06 \leq z \leq 2.35$ ) from Szomoru et al. (2013) and 81 ETGs ( $1.28 \leq z \leq 1.50$ ) from McLure et al. (2013). Sizes of data sets 4 – 8 are 168, 218, 206, 391 and 29 respectively, which consist of massive ellipticals, i.e. mass  $> 10^8 M_\odot$ . We have linearized RA and DEC for data sets 4 – 8 in the similar manner.

## 3 Method

### 3.1 Linearization of angular data

The method of cross-correlation involves only a particular type of data, spatial or linear. In earlier works, for spatial data, ‘angular distance’ between the

astronomical objects are considered whereas Euclidean distance is considered for linear data. Here the present data sets involve both spatial (RA, DEC) as well as linear (mass, size) parameters. So it is important to transform the angular data into linear data by some transformation, if any. Chattopadhyay et al. (2015) have introduced such transformation in the following manner. If the angular data  $\theta$  has the unique mode  $\phi$ , then the linear version of  $\theta$  is given by  $1 - \cos(\theta - \phi)$ . If  $\theta$  has two modes  $\phi_1$  and  $\phi_2$ , then the linear form of  $\theta$  is given by  $\max\{1 - \cos(\theta - \phi_1), 1 - \cos(\theta - \phi_2)\}$ . Mode is obtained from the angular data (here RA and DEC) by sort of circular plot of the data, where the circumference of the circle is split into groups specified by bins and the radii of the corresponding sectors are computed as equal to the square root of the relative frequencies of observations in each group. For example, Fig.1 shows the circular distribution plot of RA for 70 nearby ETGs, which is a bimodal distribution with  $\phi_1 \sim 45^\circ, \phi_2 \sim 180^\circ$ . Similarly Fig.2 shows the semi-circular distribution plot of Dec for 70 nearby ETGs, which is a unimodal with  $\phi \sim -45^\circ$ .

## 3.2 Compatibility test

Since the objects in data sets 4–8 have been collected from different sources, they obviously suffer from selection biases and errors, etc. Hence compatibility test is necessary and we performed Duda-Hart test (Duda & Hart 1973) with respect to each of the four variables to check homogeneity between two data sets from different sources with common redshift region (Table1). The test is performed within the same redshift zone as the galaxies have undergone cosmological evolution via merger and accretion (Khochfar & Silk 2006; De Lucia & Blaizot 2007; Guo & White 2008; Kormendy et al. 2009; Hopkins et al. 2010; Naab 2013). Data from Damjanov et al. (2011) contains maximum number of galaxies within the entire redshift zone ( $0.5 \leq z \leq 2.7$ ). For this we have compared its various subsets with the other sets in the corresponding redshift zones. The results are given in Table 1. It is clear from Table 1 that all the tests are accepted at 5% level of significance with high or moderate p-values. Hence, we assume that the combined data sets are compatible with each other with respect to the above four parameters size, mass, linearized RA and DEC respectively.

### 3.3 Completeness test

For testing completeness of the combined data sets 4 – 8. we have used  $V/V_{max}$  test. The test was first used by Schmidt (1965, *pg* – 513) while studying the space distribution of a complete sample of radio quasars. Let  $F_m$  be the limiting flux within maximum distance  $r_m$  and  $V(r) = \frac{4\pi r^3}{3}$ ,  $V_{max} = \frac{4\pi r_m^3}{3}$ ,  $r$  being the radial distance of a quasar. If all the quasars follow a uniform distribution over the entire range of observation then  $V/V_{max}$  will be uniformly distributed over  $[0, 1]$  and then  $\langle V/V_{max} \rangle = 0.5$ . We obtain  $\langle \log R_e / \log R_{e,max} \rangle \sim 0.397$ ,  $\langle \log M / \log M_{max} \rangle \sim 0.813$ ,  $\langle \log RA / \log RA_{max} \rangle \sim 0.545$  and  $\langle \log Dec / \log Dec_{max} \rangle \sim 0.243$  (computed after removing few outliers). For parameters with negative values their absolute values have been considered and suffix ‘*max*’ represents the maximum value of the corresponding parameter. Hence we can conclude that the combined data set is at least complete up to an accuracy of 70%. For comparison we have also plotted the combined data sets 1 – 3 and 4 – 8 in the mass-size plane (Figs 3 and 4 respectively) and in the RA-DEC plane (Figs 5 and 6 respectively).

### 3.4 Cross-correlation function

The theory of cross-correlation function was first introduced by Neyman & Scott (1952) for studying galaxy clustering on the basis of four assumptions, e.g. (i) galaxies have a natural tendency to occur in groups, (ii) the number of galaxies in each group follow certain probabilistic law, (iii) the galaxies in each group also follow certain probabilistic distribution, (iv) the distribution of groups also follow certain probabilistic law. Subsequently the above theory has been discussed by several authors, e.g. Peebles (1980); Blake et al. (2006); Martinez & Saar (2002), etc. In a previous work (De et al. 2014), the theory was applied to the mass-size distribution of galaxies, which are intrinsic properties of galaxies. Thus cross-correlation in the above aspect may be termed as parametric cross-correlation. In the present work, mass-size have been combined with spatial coordinates RA and DEC which are angular data. So we have converted angular data to linear form (viz. Subsection 3.1) and termed this new correlation function as spatio-linear cross-correlation function ( $\xi$ ). The estimate of the above function  $\hat{\xi}$  is of the form (Blake et al. 2006),

$$\hat{\xi} = \frac{D_1 D_2(r) - D_1 R_2(r) - D_2 R_1(r) + R_1 R_2(r)}{R_1 R_2(r)}, \quad (1)$$

where  $D_1 D_2(r)$  is a frequency corresponding to separation  $r$  to  $r + \delta r$  for a bin width  $\delta r$  in the histogram of the distribution  $r$ .  $D_i R_j$  and  $R_i R_j$ ,  $i, j = 1, 2$  denote similar pair counts corresponding to real samples  $D_i$  ( $i = 1, 2$ ) and simulated samples  $R_i$  ( $i = 1, 2$ ) respectively.  $r$  is the distance (here normalized Euclidean distance) between a pair of points (here galaxies) in two populations ‘1’ and ‘2’, having coordinates mass-size-RA and DEC (linearized) respectively. For error estimate we compute  $\hat{\xi}$ , 100 times and the mean value with standard error is taken as the value of the estimate  $\hat{\xi}$ . Due to noisy data initially  $\hat{\xi}$  showed an absurd pattern or was not computable in different bins. To overcome this problem we use Kernel Principal Component Analysis (Schölkopf & Smola 2002) on the data sets (viz. Section 3.5). We have used Kernel Principal Components as study variables instead of original parameters and then  $\hat{\xi}$  is computed on the basis of these variables.

### 3.5 Kernel Principal Component Analysis

It is evident that data sets 1 – 8 contain values of the parameters RA, DEC linearized and mass, size in each of them. These data sets contain noise, i.e. less information compared to the true values, giving absurd pattern or lack of finite values in each bin while computing  $\hat{\xi}$ . To overcome this problem we have used Kernel Principal Component Analysis (KPCA) on these data sets. KPCA extracts the relevant nonlinear information from raw data in terms of Kernel Principal Components (KPCs). KPCA is used for feature extraction, dimension reduction, classification, clustering, noise reduction, pattern recognition, etc. It has been successfully applied to supernovae clustering (Ishida & de Souza 2013, 2012), image denoising (Rasmussen et al. 2012), clustering of gamma-ray bursts (Modak et al. 2016), denoising of chaotic time series (Jade et al. 2003), etc. In the present context KPCA has been used for noise reduction in the raw data sets and hence to draw actual information underlying the data sets. For  $N$  parameters having  $M$  observations of each one we obtain  $M$  nonlinear components for each of  $N$  parameters by some nonlinear transformation on  $N$  parameters, which eventually concerns some matrix  $K$  called the kernel matrix. From the matrix  $K$  we obtain  $M$  eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M \geq 0$  and  $M$  eigenvectors, and associating them we get  $M$  KPCs. From these  $M$  components we select very few components  $k$  (say)  $\ll M$  for significant reduction of noise and more relevant information by trial and error method. In the present case  $k = 4$  gives the most satisfactory cross-correlation function. The details of the method has

been given in the Appendix-A.

### 3.6 Different steps in simulation and computation

To compute the cross-correlation function, given the observed sample ( $D_i$ ), we have generated random sample ( $R_i$ ) and proceeded through the following steps.

(i) A pair of observed quadruplicate (size, mass, linearized RA, linearized DEC) samples,  $D_1$  and  $D_2$  are selected. Let their sample sizes are  $n_1$  and  $n_2$  respectively.

(ii) By KPCA the first few KPCs (see Appendix-A) of  $D_1$  and  $D_2$  are found, denoted by  $D'_1$  and  $D'_2$  respectively. So our pair of study samples are now  $D'_1$  and  $D'_2$ .

(iii) For each study variable, random sample with replacement of size  $n_1$  is drawn from the set of values of the corresponding variable in  $D'_1$ . Thus we form  $n_1$  unclustered samples from  $D'_1$ , denoted by  $R_1$ , by allocating the variable-wise random samples in quadruplicate random vector. Similarly, we obtain  $n_2$  unclustered samples from  $D'_2$ , denoted by  $R_2$ .

(iv) We compute the counts in each bin along the normalized Euclidean distances calculated between  $D'_1$  and  $R_2$ ,  $D'_2$  and  $R_1$ , and  $R_1$  and  $R_2$  respectively.

(v) We repeat steps (iii) – (iv) 100 times i.e., draw  $R_1$ ,  $R_2$  and compute the counts 100 times and the mean values of the counts give us distances  $D'_1 R_2(r)$ ,  $D'_2 R_1(r)$ , and  $R_1 R_2(r)$  respectively.

(vi) Compute  $\hat{\xi}$  using formula (1), where  $D_1$  and  $D_2$  are replaced by  $D'_1$  and  $D'_2$  respectively.

(vii) Now we compute  $\hat{\xi}$  100 times. The mean value and the standard error of  $\hat{\xi}$  finally give the estimator of CCF and the standard error of the estimator of CCF respectively.

(viii) Steps (i) – (vii) are repeated considering different pair of data sets.

## 4 Results and discussion

We have computed the cross-correlation function for each of the data sets 1–3 with data sets 4–8 through KPCA, i.e. we have explored that whether there is any correlation between the population of nearby massive ETGs with high



redshift ETGs in five redshift bins as discussed in Section 2. We have found significant correlation between data set 1 and data set 8 (Fig.7), where study variables are the first four KPCs extracted from the noisy data sets using kernel (5) with  $p = 0.1$ . Moderate correlation is found between data set 2 and data set 4 (Fig.8) based on the first four KPCs extracted through (5) with  $p = 0.5$  from both the data sets. Study with the first four KPCs extracted through (5) with  $p = 1.5$  and  $p = 0.2$  from data set 3 and data set 4 respectively show moderate correlation (Fig.9). The above results indicate that the inner most component of massive ellipticals are strongly correlated with the highest redshift bin of ETGs (viz.  $2.0 \leq z \leq 2.7$ ), known as ‘red nuggets’ and the ‘intermediate’ as well as the ‘outermost part’ both are correlated with ETGs in the redshift bin  $0.5 \leq z \leq 0.75$  (Figs.7-9). The above results are very much consistent with the two phase formation model of massive nearby ETGs as suggested by various authors (Khochfar & Silk 2006; De Lucia & Blaizot 2007; Guo & White 2008; Kormendy et al. 2009; Hopkins et al. 2010; Naab 2013). The elegance of the work owes its origin from considering both the spatial as well as other intrinsic parameters, e.g. mass and effective radii of galaxies, i.e. the galaxies are not only clustered with respect to spatial coordinates but also with respect to other intrinsic parameters of the galaxies, thus describing a complete scenario for the formation of massive ETGs.

After estimating the cross-correlation function between data sets 1 and 8, 2 and 4, 3 and 4, we have fitted a power law of the form

$$\xi(r) \propto \frac{1}{r}, \quad (2)$$

i.e.

$$\xi(r) = Ar^{-1}, \quad (3)$$

for each of the functions, where the values of  $A$  are given in Figs.7-9. We have also performed Kolmogorov-Smirnov test (Kolmogorov 1993) for testing goodness of fit. The power law fitting has a great significance in the sense that it helps us to reject a model with no correlation and the asymptotic nature as  $r \rightarrow 0$ . This indicates a stronger correlation for smaller values of  $r$ , i.e. the clustering nature of ETGs is stronger for lesser distance both in spatio-linear aspect. That is, the galaxies are not only clustered in the spatial sense but also clustered with respect to their linear intrinsic properties. The  $p$ -values are also listed in Figs.7-9.

We have already discussed that discovery of ‘red nuggets’ at high redshift has challenged the formation of massive ellipticals as a ‘single event’, e.g.

‘monolithic’ collapse or ‘major merger’. Nowadays ellipticals having maximum masses undergo through a number of steps. They are formed at  $z \sim 6$  or higher through a dissipative process and subsequently become very massive  $\sim (10^{11} M_\odot)$  and compact ( $R_e \sim 1 \text{ kpc}$ ) in a very short interval of time  $z \sim 2$  (Dekel et al. 2009; Oser et al. 2010; Oser et al. 2012). In spite of the above event a significant fraction remains less active at  $z \sim 2$ . They are 4 – 5 times less compact and less massive by a factor of 2 corresponding to their low-redshift descendants (Buitrago et al. 2008; van Dokkum 2008; van der Wel et al. 2008; Cimatti et al. 2008; Bezanson et al. 2009; van Dokkum et al. 2010; Whitaker et al. 2012).

For the massive ellipticals in the present sample, the innermost cores (data set 1) is strongly correlated with galaxies in the highest redshift zone ( $2.0 \leq z \leq 2.7$ ) and their core masses have the median values  $\sim 10^{10} M_\odot$  and  $\sim 10^{11} M_\odot$  respectively. Hence we can conclude reasonably that these high redshift population forms the cores of at least some, if not all (Graham et al. 2015; Wellons et al. 2016), present day massive ellipticals. Thus massive ellipticals can not be formed by only monolithic collapse, otherwise they will be too small and too red (van Dokkum 2008; Ferré-Mateu et al. 2012). Regarding the formation of intermediate and outer part (data sets 2 – 3) it might be explained as a result of major or minor mergers following Naab, Johansson & Ostriker (2009). According to the above authors if  $M_i, r_i, E_i$  and  $\langle v_i^2 \rangle$  be the initial mass, radius, energy and mean squared speed of a stellar system and  $M_a, r_a, E_a$  and  $\langle v_a^2 \rangle$  be the corresponding values after a merger with other systems then,

$$\frac{\langle v_f^2 \rangle}{\langle v_i^2 \rangle} = \frac{(1 + \eta\epsilon)}{1 + \eta},$$

$$\frac{r_{g,f}}{r_{g,i}} = \frac{(1 + \eta)^2}{1 + \eta\epsilon},$$

$$\frac{g_f}{g_i} = \frac{(1 + \eta\epsilon)^3}{(1 + \eta)^5},$$

where the quantities with suffix ‘ $f$ ’ denote the final values,  $\eta = M_a/M_i$ ,  $\epsilon = \langle v_a^2 \rangle / \langle v_i^2 \rangle$ ,  $g$  is the density and ‘ $r'_g$ ’ stands for gravitational radius of the system. When  $\eta = 1$  it implies a major merger and then the size increases by a factor 2. In our case the intermediate part (viz. data set2) has mean radius  $\langle R_{e,2} \rangle \sim 2.276 \text{ kpc}$  which is almost 7.689 times larger than the

mean radii of innermost part,  $\langle R_{e,1} \rangle \sim 0.296$  kpc. Hence we can conclude that the intermediate parts of nearby massive ellipticals have been formed by major mergers. The result is also consistent with our previous works (Chattopadhyay et al. 2009; Chattopadhyay et al. 2013).

In the limiting case when  $\langle v_a^2 \rangle \ll \langle v_i^2 \rangle$  or  $\epsilon \ll 1$ , the size increases by a factor 4 and this criterion implies minor merger. In our situation  $\langle R_{e,3} \rangle \sim 19.554$  kpc and  $\langle R_{e,3} \rangle - \langle R_{e,1} \rangle \sim 19.258$  kpc which is much larger than the average value of the innermost part. Also the median mass is of the order of  $10^{11} M_\odot$  which is comparable to the combined masses of few dwarf galaxies. This indicates that the outermost part of nearby massive ETGs might have been formed by several minor mergers and at the same time strong correlation exists between outer parts of these galaxies with the highest redshift ETGs in the range  $0.5 < z \leq 0.75$ , i.e. the same as that for the intermediate part. The present result is more authentic support in a quantitative manner for a second phase for the formation of present massive ETGs and the treatment of considering both parameters spatial as well as intrinsic, helps us to obtain a more robust picture for the two-phase formation of massive ellipticals. It overcomes the controversy raised in a previous work (De et al. 2014) for the formation of the outermost parts of these galaxies. The above result is also consistent with the result obtained by Mondal et al. (2008) where halos of massive galaxies are found to be formed by tidal stripping of satellite dwarf galaxies.

The present work also supposed to encourage future studies with mixed type data.

## A Kernel principal component analysis (KPCA)

Let there be  $M$  observations  $x_k, k = 1, 2, \dots, M$  of  $N$  parameters, i.e.  $x_k \in R^N$ . It is assumed that  $\sum_{k=1}^M x_k = 0$ . If the condition does not hold, there exists transformation which converts  $x_k$  to the above form. KPCA is performed on a dot product space  $F$  (feature space) by using a map from  $R^N$  to  $F$  defined as,

$$\Phi : R^N \rightarrow F.$$

We need to solve the following equation for eigenvalues  $\lambda \geq 0$ , and eigenvectors  $V \in F$  (non-zero vector)

$$\lambda V = \bar{C}V, \text{ where } \bar{C} = \frac{1}{M} \sum_{j=1}^M \Phi(x_j)\Phi(x_j)^T.$$

The above relation actually leads to solving the eigen value problem for the kernel matrix  $K = ((K_{ij}))_{i,j=1(1)M}$  as follows:

$$M\lambda\alpha = K\alpha,$$

where  $K_{ij} = \langle \Phi(x_i), \Phi(x_j) \rangle$ ,  $\langle \cdot, \cdot \rangle$  represents the usual dot product, is a positive definite (pd) kernel.  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M (\geq 0)$  denote the eigenvalues of  $K$  and  $\alpha^1, \dots, \alpha^M$  be the corresponding complete set of eigenvectors, with  $\lambda_l$  being the last nonzero eigenvalue. Then the  $k^{th}$  kernel principal component (KPC) corresponding to  $\Phi(x)$  is given by

$$\langle V^k, \Phi(x) \rangle = \sum_{i=1}^M \alpha_i^k \langle \Phi(x_i), \Phi(x) \rangle = \sum_{i=1}^M \alpha_i^k k(x_i, x), \quad k = 1, 2, \dots, l, \quad (4)$$

where  $k(x_i, x)$  = kernel corresponding to  $x_i$  and  $x$ .

Now the assumption  $\sum_{k=1}^M x_k = 0$  can be relaxed by using the kernel matrix  $\tilde{K}$  in the place of  $K$ , where  $\tilde{K}_{ij} = (K - 1_M K - K 1_M + 1_M K 1_M)_{ij}$ ,  $(1_M)_{ij} = M^{-1}$ , for  $i, j = 1, 2, \dots, M$ . Then the kernel in KPCA can be a conditionally positive definite (cpd) kernel, while ‘cpd’ kernels include ‘pd’ kernels (Hofmann et al. 2008). In our study we use the following kernel (Modak et al. 2016)

$$k(x, y) = \exp\left(-\sum_{i=1}^N \left|\frac{x_i - y_i}{s_i}\right|^p\right), \quad (5)$$

between  $x, y \in R^N$ , where  $p$  ( $0 < p \leq 2$ ) is a tuning parameter and  $s_i$  ( $> 0$ ),  $i = 1, 2, \dots, N$  are scale parameters. It is a symmetric and ‘pd’ kernel.

The higher the order of KPC, the less relevant information and the more noise are supposed to be contained in that component (Schölkopf & Smola 2002). Hence noise can be effectively reduced by taking into account only first few KPCs and discarding the others, which is also shown in terms of pre-images in Schölkopf & Smola 2002. So we start from the first two KPCs

and take up to the first four KPCs (not exceeding four means not to increase the dimension of the data sets) as the study variables. We take  $s_i$ =sample standard deviation of the  $i^{th}$  KPC and vary  $p$  in a trial-and-error method. We choose the number of KPCs and the value of  $p$  such that we can extract important features from the data sets separating out the noise in the form of nonlinear KPCs and hence get the most satisfactory results.

## Acknowledgments

Authors (T.C & A.K.C) are grateful for getting partial support from Indo-French project (Project No. 15EP06, 2015-17) for the work.

## References

- [1] Arimoto, N., & Yoshii, Y. 1987, A&A, 173, 23
- [2] Ashman, K. M., & Zepf, S. E. 1992, ApJ, 384, 50
- [3] Bernardi, M., Roche, N., Shankar, F., & Sheth, R. K. 2011, MNRAS, 412, L6
- [4] Bezanson, R., van Dokkum, P. G., Tal, T., et al. 2009, ApJ, 697, 1290
- [5] Blake, C., Pope, A., Scott, D., & Mobasher, B. 2006, MNRAS, 368, 732
- [6] Bluck, A. F. L., Conselice, C. J., & Buitrago F. 2012, ApJ, 747, 34
- [7] Bok, B. J. 1934, Bull. Harv. Obser., 895, 1
- [8] Buitrago, F., Trujillo, I., & Conselice, C. J. 2008, ApJ, 687, L61
- [9] Cappellari, M., di Serego Alighieri, S., Cimatti, A., et al. 2009, ApJ, 704, L34
- [10] Carlberg, R. G. 1984, ApJ, 286, 403
- [11] Chattopadhyay, A. K., Chattopadhyay, T., Davoust, E., Mondal, S., & Sharina, M. 2009, ApJ, 705, 1533

- [12] Chattopadhyay, A. K., Mondal, S., & Chattopadhyay, T. 2013, CSDA, 57, 17.
- [13] Chattopadhyay, A. K., Mondal, S., & Biswas, A. 2015, Environ Ecol Stat, 22, 33
- [14] Chandrasekhar, S., & Munch, G. 1952, ApJ, 115, 103
- [15] Chen, Z., Shu, C. G., & Burgarella, D. 2013, MNRAS, 431, 2080
- [16] Cimatti, A., Cassata, P., Pozzetti, L., et al. 2008, A&A, 482, 21
- [17] Daddi, E., Renzini, A., & Pirzkal, N., 2005, ApJ, 626, 680
- [18] Damjanov I., Abraham, R. G., Glazebrook, K., et al. 2011, ApJ, 739, L44
- [19] de la Rosa, I. G., La Barbera, F., Ferreras I., et al. 2016, MNRAS, 457, 1916
- [20] Dekel, A., Sari, R., & Ceverino, D. 2009, ApJ, 703, 785
- [21] De Lucia, G., & Blaizot, J. 2007, MNRAS, 375,2
- [22] De, T., Chattopadhyay, T., & Chattopadhyay, A.K. 2014, PASA, 31, 47
- [23] Duda, R. O., & Hart, P. E. 1973, Pattern Classification and Scene Analysis (New York: Wiley)
- [24] Ferré-Mateu, A., Vazdekis, A., Trujillo, I., et al. 2012, MNRAS, 423, 632
- [25] Forbes, D.A., Spitler, L.R., Romanowsky, A.J., Brodie, J.P. & Foster, C. 2011, MNRAS, 413, 2943
- [26] Forbes, D.A., Bordie, J.P., & Grillmair, C.J. 1997, ApJ, 113,1652
- [27] Graham, A.W., Dullo, B.T., & Savorgnam, G.A.D. 2015, ApJ, 804, 32
- [28] Gobat, R., Daddi, E., Onodera, M., et al. 2011, A&A, 526, A133
- [29] Guo, Q., & White, S. D. M. 2008, MNRAS, 384, 2

- [30] Ho, L. C., Li, Z.-Y., Barth, A. J., Seigar, M. S., & Peng, C. Y. 2011, *ApJS*, 197,21
- [31] Hofmann, T., Schölkopf, B., & Smola A. 2008, *Ann. Statist.*,36,1171
- [32] Hopkins, P. F., Croton, D., & Bundy, K. 2010, *ApJ*, 724, 915
- [33] Huang, S., Ho, L. C., Peng, C. Y., Li, Z.-Y., & Barth, A. J. 2013, *ApJ*, 768, L28
- [34] Ishida, E. E. O., 2012, *Proceedings of the International Astronomical Union*, 1, 683
- [35] Ishida, E. E. O., & de Souza, R. S. 2013, *MNRAS*, 430, 509
- [36] Jade, A. M., Srikantha, B., Jayaramana, V. K., et al. 2003, *Chemical Engineering Science*, 58, 4441
- [37] Johansson, P. H., Naab, T., & Ostriker, J. P. 2012, *ApJ*, 754, 115
- [38] Kaufman, L., & Rousseeuw, P. J. 1990, *Finding Groups in Data: An Introduction to Cluster Analysis* (New York: Wiley)
- [39] Khochfar, S., & Silk, J., 2006, *ApJ*, 648, L21
- [40] Kolmogorov, A., 1993, *Ist. Ital. Attuari*, 4, 834
- [41] Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, *ApJS*, 182, 216
- [42] Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
- [43] Larson, R. B. 1975, *MNRAS*, 173, 671
- [44] Limber, D. N. 1953, *ApJ*, 117, 134
- [45] Limber, D. N. 1954, *ApJ*, 119, 655
- [46] Martínez, V. J., & Saar, E. 2002, *Statistics of the Galaxy Distribution* (Boca Raton: Chapman & Hall/CRC)
- [47] McLure, R. J., Pearce, H. J., Dunlop, J. S., et al. 2013, *MNRAS*, 428, 1088

- [48] Modak, S., Chattopadhyay, A. K., & Chattopadhyay, T., 2016, AJ, submitted
- [49] Mondal, S., Chattopadhyay, T., & Chattopadhyay, A. K. 2008, ApJ, 683, 172
- [50] Mowbray, A. G. 1938, PASP, 50, 275
- [51] Naab, T., Johansson, P. H., & Ostriker, J. P., 2009, ApJ, 699, L178
- [52] Naab, T., 2013, Modelling the formation of today's massive ellipticals. The Intriguing Life of Massive Galaxies. Proceedings of the International Astronomical Union, IAU Symposium, 295, 340
- [53] Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, ApJ, 746, 162
- [54] Neyman, J., & Scott, E. L. 1952, ApJ, 116, 144
- [55] Neyman, J., Scott, E. L., & Shane, C. D., 1954, ApJS, 1, 269
- [56] Onodera, M., Renzini, A., Carollo, M., et al. 2012, ApJ, 755, 26
- [57] Oser, L., Naab, T., Ostriker, J. P., & Johansson, P. H., 2012, ApJ, 744, 63
- [58] Oser, L., Ostriker, J. P., Naab, T., Johansson, P. H., & Burkert, A., 2010, ApJ, 725, 2312
- [59] Peebles, P. J. E., 1980, The Large-scale Structure of the Universe (Princeton, NJ: Princeton University Press)
- [60] Prieto, M., Eliche-Moral, M. C., Balcells, M., et al. 2013, MNRAS, 428, 999
- [61] Papovich C., Bassett, R., Lotz, J. M., et al. 2012, ApJ, 750, 93
- [62] Rasmussen P. M., Abrahamsen T. J., Madsen K. H., Hansen L. K., 2012, NeuroImage, 60, 1807
- [63] Schmidt, M. 1965, in Star and Stellar Systems, Galactic Structures, Vol.5 (ed. A. Blaauw, & M. Schmidt; Chicago: University of Chicago Press)



- [64] Schölkopf, B., & Smola, A., 2002, Learning with kernels: Support vector machines, regularization, optimization, and beyond (Cambridge,Massachusetts:MIT Press)
- [65] Szomoru, D., Franx, M., & van Dokkum, P. G. 2012, ApJ, 749, 121
- [66] Szomoru, D., Franx, M., van Dokkum, P. G., et al. 2013, ApJ, 763, 73
- [67] Toomre, A., & Toomre, J. 1972, 178, 623
- [68] van der Wel, A., Rix, H.-W., Wuyts, S., et al. 2011, ApJ, 730, 38
- [69] van Dokkum, P. G. 2008, ApJ, 674, 29
- [70] van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 709, 1018
- [71] van Dokkum, P.G., Nelson, E. J., Franx, M., et al. 2015, ApJ, 813, 23
- [72] Vulcani,B., Marchesini, D., De Lucia, G., et al. 2016, ApJ, 816, 86
- [73] Wellons, S., Torrey, P., Ma, C.P., et al. 2016, MNRAS, 456, 1030
- [74] Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJ, 754, L29
- [75] Zepf, S. E., Beasley, M. A., Bridges, T. J.,et al. 2000, AJ, 120, 2928
- [76] Zwicky, F. 1953, AcHPh, 26, 241

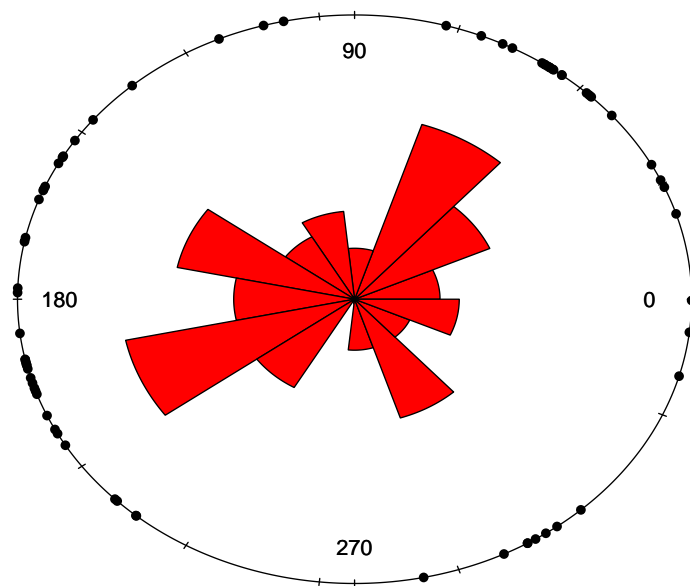


Figure 1: Distribution plot of right ascension (RA) for 70 nearby ETGs (Huang et al. 2013).

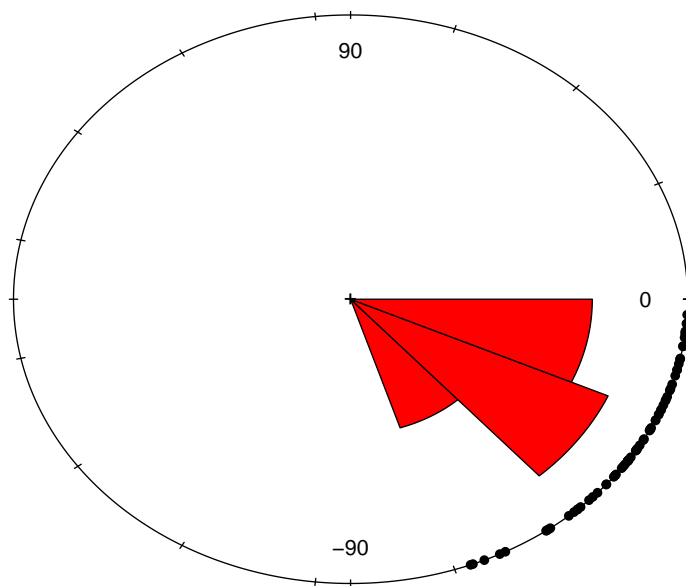


Figure 2: Distribution plot of declination (DEC) for 70 nearby ETGs (Huang et al. 2013).

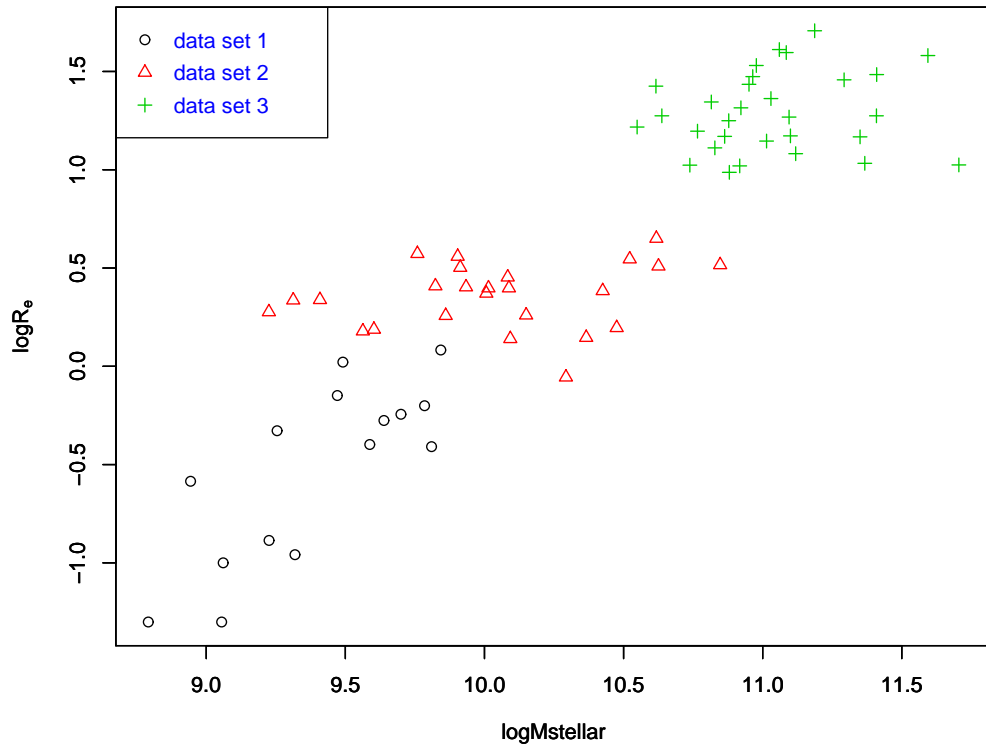


Figure 3:  $\log M$  versus  $\log R_e$  plot of the data points for the data sets 1 – 3.

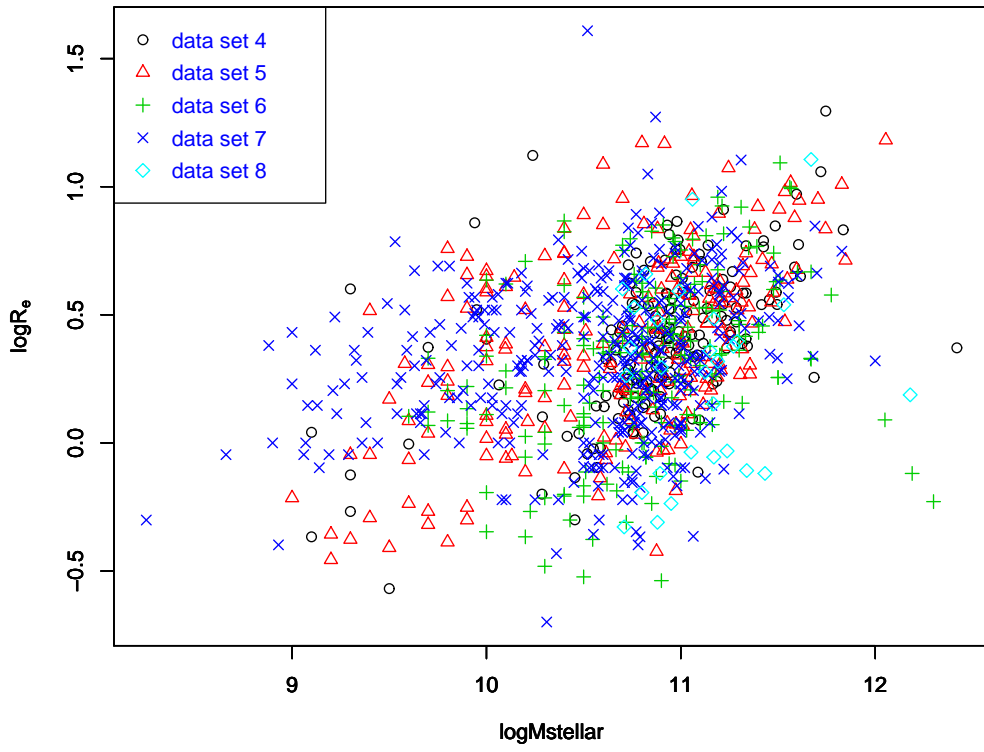


Figure 4:  $\log M$  versus  $\log R_e$  plot of the data points for the data sets 4 – 8.

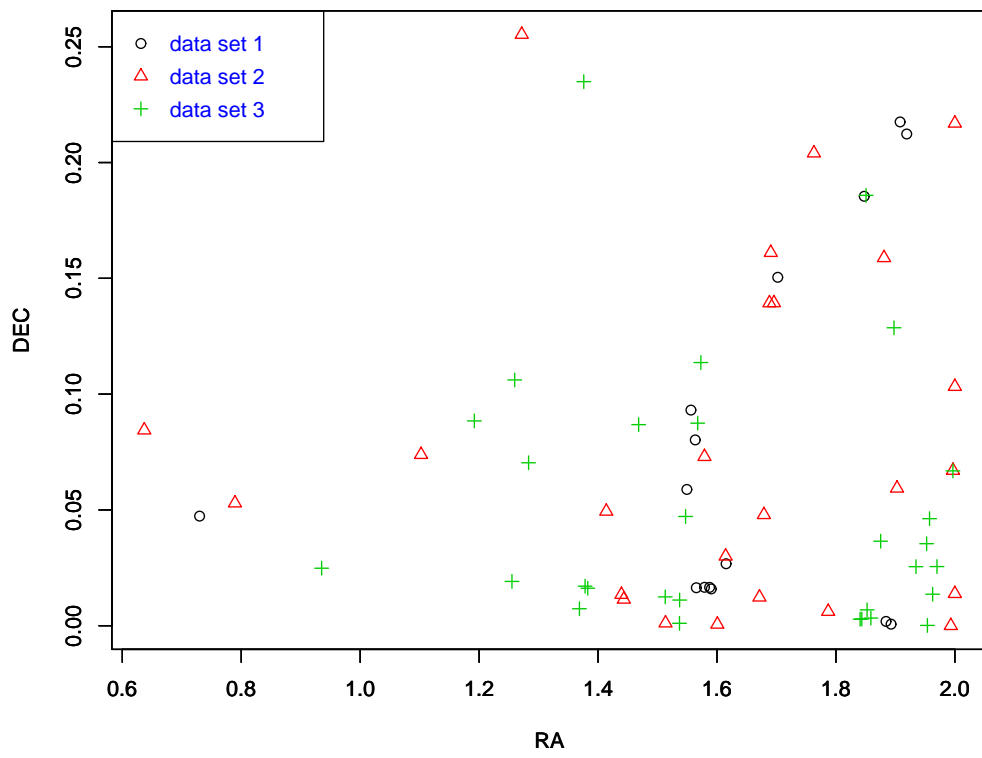


Figure 5: Linearized  $RA$  versus  $DEC$  plot of the data points for the data sets 1 – 3.

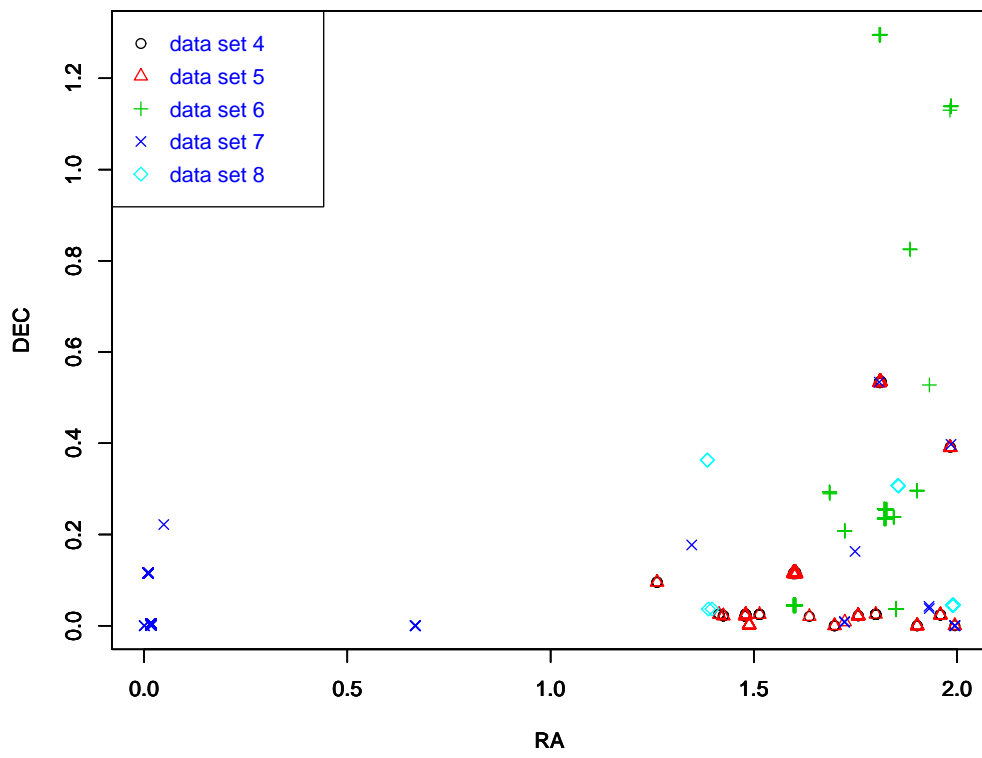


Figure 6: Linearized  $RA$  versus  $DEC$  plot of the data points for the data sets 4 – 8.

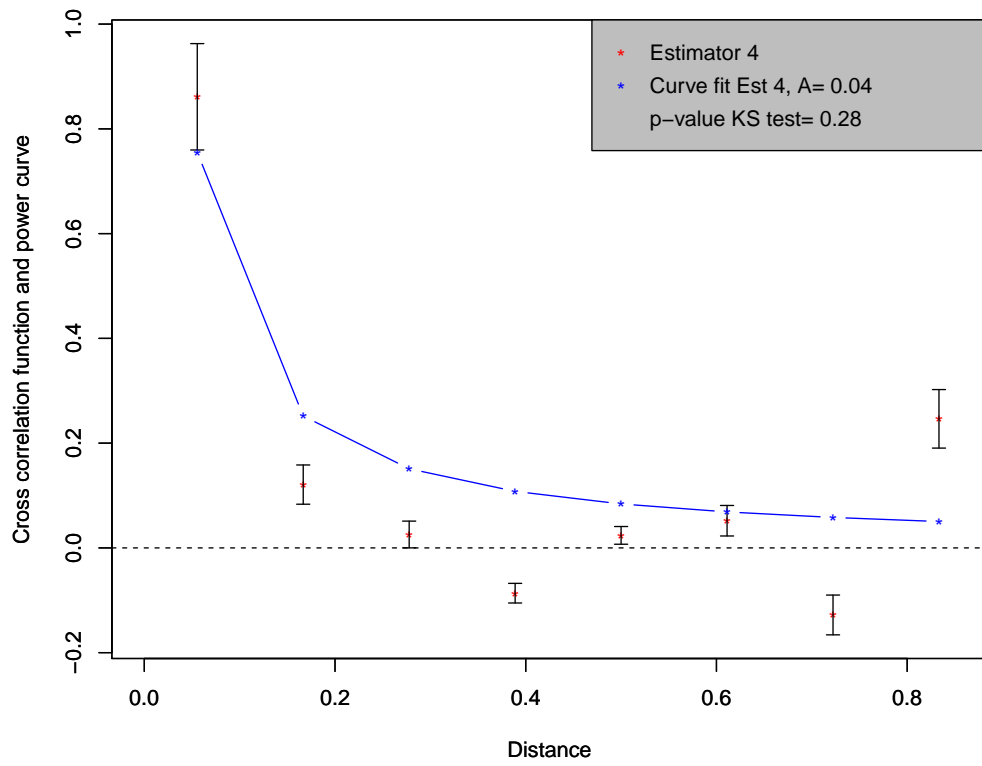


Figure 7: Cross correlation function (CCF) plot for data sets 1 and 8, solid line represents fitted power law  $\xi(r) = Ar^{-1}$  to the CCF estimator.



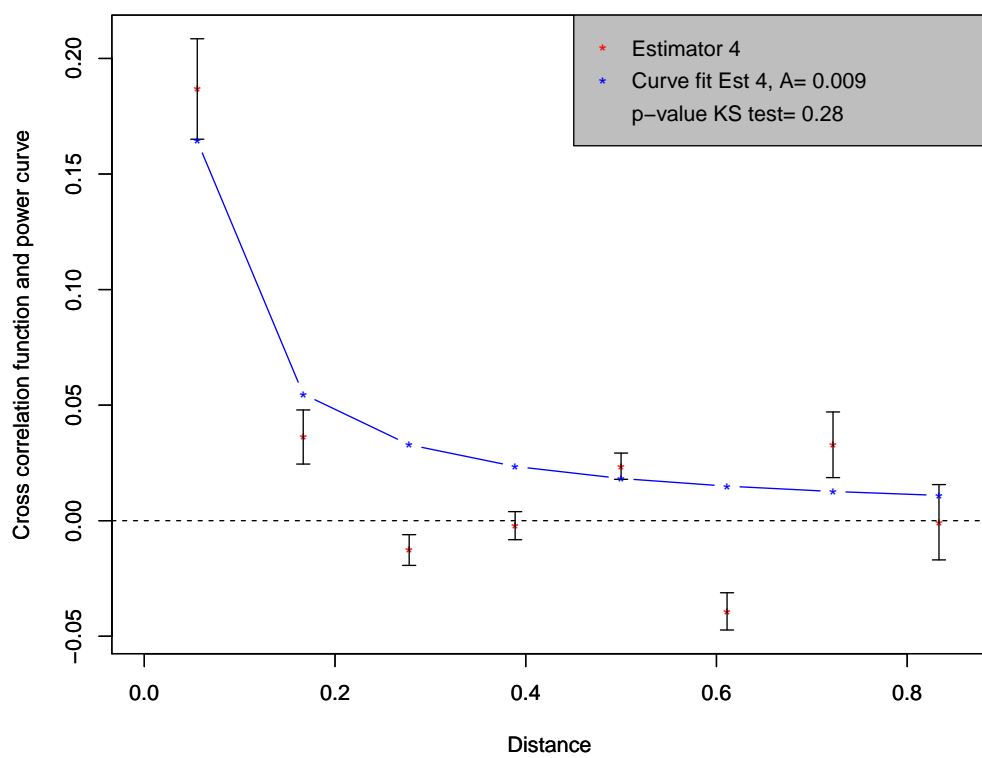


Figure 8: Same as in Fig. 7 but for data sets 2 and 4.

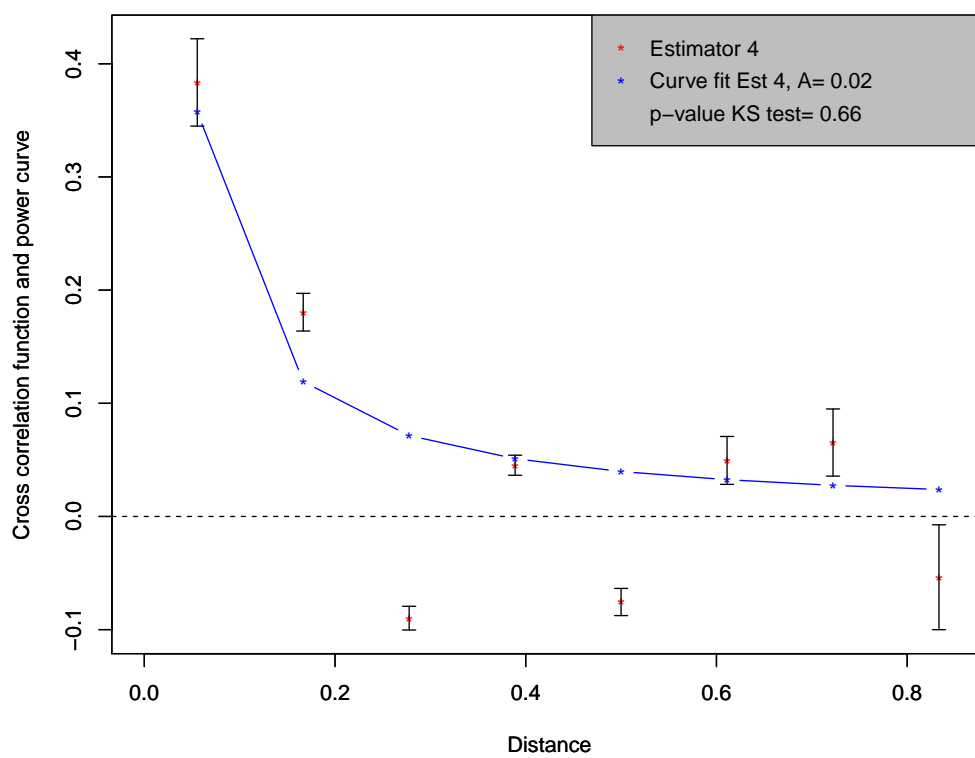


Figure 9: Same as in Fig. 7 but for data sets 3 and 4.

Table 1: Homogeneity test at 5% level of significance for size, mass, linearized RA and DEC among various data sets in the corresponding redshift zones.

Sample <sub>1</sub>	Sample <sub>2</sub>	$z$	no. of ETGs in 2 samples	p-value variable-wise	Decision
Damjanov et al. (2011)	Papovich et al. (2012)	[1.49, 1.79]	(22, 248)	(1, 1, 0.413, 1)	Accept
”	Chen et al. (2013)	[0.703, 1.309]	(199, 142)	(1, 0.998, 1, 1)	Accept
”	Szomoru et al. (2013)	(0.5,2.7)	(364, 177)	(1, 1, 1, 1)	Accept
”	McLure et al. (2013)	[ 1.28, 1.505]	(36, 81)	(1, 1, 0.389, 1)	Accept